

**Apparatus and Method for Prediction and Management of
Subject Compliance in Clinical Research**

Reference to Related Applications

- 5 The subject matter of this application relates to the patent application titled “System for Clinical Trial Subject Compliance”, Attorney Docket No. IVQ-002, and filed on even date herewith. The aforementioned application, and the references cited therein, are incorporated herein by reference.

10 **Field of the Invention**

 The present invention relates to predicting subject behavior during research, especially clinical trials. Specifically, the invention relates to the prediction of subject noncompliance with protocols in clinical trials.

15 **Background of the Invention**

- Evaluation of subject compliance with research protocols typically looks at only one variable at a time. Such evaluation is not empirically derived by quantitative analysis of existing datasets, instead relying on the researcher’s judgment and biases to determine whether and what type of corrective action is required. Furthermore,
- 20 evaluation of subject compliance with clinical trial protocols has typically not taken into account the domain of the clinical trial or the characteristics of the subjects. Finally, such evaluation often cannot be made in a timely way, but is made only after serious noncompliance has already occurred.

25 **Summary of the Invention**

- The goal of clinical trials is to collect valid, reliable data on one or more conditions within a clinical trial group of subjects. Subjects in clinical trials are assigned tasks related to treatment and data collection in accordance with a research protocol. The integrity of clinical trials rests upon subjects’ faithful performance of
- 30 these tasks. If subjects fail to comply with the protocol, the trial fails to yield reliable, valid results. Thus, subject noncompliance in clinical trials is a significant risk and cost to the pharmaceutical industry. Accordingly, predicting subject performance and assessment of such performance is of substantial value to clinical research.

The benefits of a system that can predict and track subject compliance in a clinical trial include: reliable, valid data; increased statistical power; reduced clinical trial costs through smaller sample sizes; reduced time to complete the clinical trial; and, ultimately, reduced time to get a drug or medical device to market.

5 According to one embodiment of the invention, a method of predicting subject noncompliance is provided. The method includes the steps of providing historical subject compliance data, generating at least one predictive algorithm for predicting subject noncompliance by quantitative analysis of the historical subject compliance data, and translating the at least one predictive algorithm into at least one prediction
10 rule for use with a clinical trial.

According to another embodiment, a method of determining subject noncompliance includes the steps of providing at least one of the group of historical subject compliance data and historical protocol data and generating at least one algorithm reflective of at least one of historical subject compliance data and historical
15 protocol data by quantitatively analyzing the historical subject compliance data and the historical protocol data. The method also includes translating the algorithm into at least one decision rule for analyzing subject compliance information, obtaining the subject compliance information and comparing the subject compliance information to the at least one decision rule to determine if corrective action is needed.

20 According to a further embodiment, a method of the invention includes the steps of providing historical subject compliance data and historical protocol data, generating a spectrum of noncompliance representative of the historical subject compliance data not compliant with the historical protocol data by quantitative analysis of the historical subject compliance data and the historical protocol data,
25 obtaining subject compliance information and comparing the spectrum of noncompliance to the subject compliance information to determine if corrective action is needed.

According to an embodiment of the invention a method of detecting subject fraud is provided, having the steps of providing historical subject compliance data and
30 historical protocol data, generating at least one fraud detection algorithm for detecting subject fraud by quantitative analysis of the historical subject compliance data and the historical protocol data and translating the at least one fraud detection algorithm into at least one fraud detection rule for use with a clinical trial.

According to an embodiment of the invention another method of detecting subject fraud is provided, having the steps of providing subject compliance data, generating at least one fraud detection algorithm for detecting subject fraud by quantitative analysis of the compliance data and translating the at least one fraud
 5 detection algorithm into at least one fraud detection rule for use with a clinical trial.

According to an embodiment of the invention a medium suitable for use in an electronic device and having instructions for execution on the electronic device, the instructions, having the steps of providing at least one of the group of historical subject compliance data and historical protocol data, generating at least one predictive
 10 algorithm for predicting subject noncompliance by quantitative analysis of at least one of the group of the historical subject compliance data and the historical protocol data and translating the at least one predictive algorithm into at least one prediction rule for use with a clinical trial.

According to another embodiment of the invention a medium suitable for use
 15 in an electronic device and having instructions for execution on the electronic device, the instructions having the steps of providing at least one of the group of historical subject compliance data and historical protocol data, generating at least one algorithm reflective of at least one of the group of the historical subject compliance data and the historical protocol data by quantitative analysis of the historical subject compliance
 20 data and the historical protocol data, translating the at least one algorithm into at least one decision rule for analyzing subject compliance information, obtaining the subject compliance information and comparing the subject compliance information to the at least one decision rule to determine if corrective action is needed.

According to another embodiment of the invention a medium suitable for use
 25 in an electronic device and having instructions for execution on the electronic device, the instructions having the steps of providing historical subject compliance data and historical protocol data, generating a spectrum of noncompliance representative of the historical subject compliance data not compliant with the historical protocol data by quantitative analysis of the historical subject compliance data and the historical
 30 protocol data, obtaining subject compliance information and comparing the spectrum of noncompliance to the subject compliance information to determine if corrective action is needed.

According to a further embodiment of the invention a medium suitable for use in an electronic device and having instructions for execution on the electronic device, the instructions having the steps of providing historical subject compliance data and historical protocol data, generating at least one fraud detection algorithm for detecting subject fraud by quantitative analysis of the historical subject compliance data and the historical protocol data and translating the at least one fraud detection algorithm into at least one fraud detection rule for use with a clinical trial.

According to an embodiment of the invention a medium suitable for use in an electronic device and having instructions for execution on the electronic device, the instructions having the steps of providing subject compliance data, generating at least one fraud detection algorithm for detecting subject fraud by quantitative analysis of the compliance data and translating the at least one fraud detection algorithm into at least one fraud detection rule for use with a clinical trial.

15 Brief Description of the Drawings

The foregoing and other objects, features and advantages of the invention will be apparent from the following description and apparent from the accompanying drawings.

Figure 1 illustrates a system for determining subject noncompliance according to the teachings of the present invention; and

Figure 2 is a schematic flow chart diagram illustrating the method according to the teachings of the present invention.

Detailed Description of the Invention

The present invention involves an empirically derived set of algorithms and decision rules to predict subject compliance, and detect noncompliance, with research protocols. The present invention uses algorithms and decision rules to provide an empirical approach for predicting different types of subject noncompliance with research protocols. This actuarial approach to predicting subject noncompliance with clinical trial protocols is consistent with empirical research demonstrating the superiority of actuarial prediction of human behavior as compared to subjective

clinical judgment. According to an alternative embodiment of the invention, a portable electronic device is used to query and collect data from the subject.

As used herein "clinical trial" refers to a broad range of data collecting activities, including studies directed to monitoring of one or more conditions within a clinical trial group of subjects. One such example includes drug trials involving humans. As used herein "subject" refers to any participant in a clinical trial, whether or not the subject has any relationship to a doctor or other medical professional.

"Trial data" or "clinical trial data" refers to data gathered for the principle purpose of the clinical trial. For example, trial data would include pain levels experienced by subjects in a pain medication clinical trial or craving levels in an anti-smoking medication clinical trial.

"Evaluability data" or "compliance data" or "compliance information" is data that relates to the circumstances under which the trial data was collected or other data pertaining to characteristics of the trial data or other evaluability data. Some examples include timeliness, consistency with other collected data, proximity of the data to an expected data range and completeness of the data.

"Historical protocol data" includes data specifying the research protocol of earlier clinical trials. Examples of historical protocol data can include questions posed to subjects, frequency of prompting of a subject during various times of the day or week, time allowed for subjects to respond to questions, requirements of subject behavior, and conditions mandating removal of a subject from certain statistical analyses or removal as participant in the clinical trial.

As used herein "portable electronic device" refers to any electronic device that can be adapted for use by a subject and/or clinical staff for viewing and/or inputting information. Preferably, the portable electronic device will also have a visual, audible or tactile alarm to gain the attention of the subject. For example, a pager having a vibration alarm may be used as a portable electronic device. Further examples include, pagers with audible alarms and/or text messaging capabilities, a laptop computer or a cell phone. Preferably, according to the invention, a portable electronic device will be a handheld computer provided with a display and a data input feature, such as a touch-sensitive screen, or buttons to enable a subject to respond to questions posed on the display or to input unsolicited information. Examples of such portable electronic devices include the Palm Pilot by Palm, Inc or Windows-based devices

running Pocket PC from Microsoft Corporation. Preferably, the portable electronic device will also be adapted to communicate with at least one other computer via a wireless connection or via a wired connection, including the use of a modem and/or a network, such as a LAN or the Internet.

- 5 According to an embodiment of the present invention, a system is provided as shown in Figure 1. A processor 10 is provided and is adapted to communicate with at least one database 20. As discussed below, the database preferably stores data related to subject compliance and associated research protocols. An input device 30 is also provided to allow the subject or another person to provide input to the processor 10.
- 10 The input device 30 may be a keyboard, a modem or other such device adapted for communication with the processor. An output device 40 is also preferably provided to receive and display information from the processor 10. Examples of output devices 40 include a printer and a monitor.

- 15 In one embodiment of the invention, a portable electronic device 50 is provided and is selectively operatively coupled to the processor 10. The portable electronic device 50 can also include a processor and may serve as an alarm, an input device, an output device, and/or a database. One example of a portable electronic device is a Palm Pilot by Palm, Inc, as described above.

- 20 According to an embodiment of the invention, a flow chart illustrating the method of the present invention is set forth in Figure 2. First, relevant subject compliance data, and associated protocol data, reflecting subject compliance with research protocols in clinical trials, is provided, step 110. Optionally, only subject compliance data may be provided, as some application of the present invention may not require knowledge of associated historical protocol for use of the subject
- 25 compliance data. For example, analysis of response times to questions may not require knowledge of the maximum permissible time for subjects to answer questions in earlier clinical trials.

- Subject compliance data and associated protocol data is preferably stored in one or more databases 20 and may be identified from earlier clinical trials and/or
- 30 earlier activities of a current clinical trial. An output of the present invention preferably includes a database to provide subject compliance data and associated protocol data for later use by the invention.

The subject compliance data and associated protocol data is preferably specific to the type of condition or medication that is the focus of the clinical trial. For example, if the clinical trial relates to a cardiovascular condition, the data preferably relates to subject compliance with protocols in cardiovascular clinical trials. Likewise, if the clinical trial relates to a cardiovascular medication, the data used in the present invention will preferably relate to subject compliance with protocols in cardiovascular clinical trials. It is also within the scope of the invention to optionally include subject compliance data and associated protocol data obtained from an earlier phase of the clinical trial into the compliance data of the present invention. Alternatively, the subject compliance data and associated protocol data may not be related to the type of condition or medication that is the focus of the clinical trial.

Next, at least one algorithm representative of the subject compliance data is generated by quantitative analysis of the compliance data, step 120. Preferably, multiple algorithms are generated. The present invention involves the application of statistical and other quantitative methods to screen existing research data for markers of, e.g. variables related to, noncompliance with research protocols. Preferably, the subject compliance data is also reviewed to exclude invalid data. For example, data reported by one subject that appears to be well outside a range established by all other subjects can indicate invalid data.

Examples of various types of data that may be collected according to an embodiment of the invention include variables that may represent 'non-intuitive' predictors such as: gender of the subject, disease severity, the time of the year, and the day of the week.

Quantitative analysis methods are used to distinguish, identify, and predict instances of good and poor compliance and/or instances of valid or invalid data entries. The quantitative analysis methods of the present invention may include, but are not limited to, application of a variety of statistical and data mining techniques, such as logistic regression, discriminant function analysis, classification and regression trees, neural networks, and multiple linear regression to screen existing data and derive algorithms to identify markers of noncompliance with research protocols.

Logistic regression analyses use dichotomous and continuous variables to predict dichotomous outcome variables. For example, dichotomous outcome variables can indicate "completed" or "failed to complete" a clinical trial monitoring protocol.

Discriminant function analysis relates a categorical criterion variable to dichotomous or linear predictors. Classification and Regression Trees (CART) use binary recursive partitioning to identify unique groups of subjects, such as, for example, subjects failing to complete the clinical trial protocol and subjects completing the protocol with minimal corrective feedback regarding their compliance with the clinical trial protocol. Neural network approaches to pattern recognition examine data for patterns and classify certain patterns of data based on the probability that they lead to a given outcome. Multivariate Regressive Splines (MARS) build flexible regression models, including interactions, by fitting separate splines to distinct intervals of the predictor variables.

Other nonparametric and parametric statistical approaches can also be applied to the prediction of subject noncompliance with clinical trial protocols.

A variety of predictor and criterion variables can be used in the present invention. Predictor variables can range between basic demographic characteristics, such as, for example, gender, to specific clinical trial compliance related variables, such as, for example, mean latency to respond to an audible prompt from an optional portable electronic device. Similarly, criterion variables can range from subtle, such as, for example, missing some percentage of portable electronic device prompts, to severe noncompliance, such as, for example, failure to complete the clinical trial protocol. For detection of fraudulent entries, example predictor variables could include the speed or rate of entries, or an implausible or statistically unlikely pattern of answers to a set of assessment questions.

The present invention allows for complex, non-intuitive interactions among multiple variables to optimally predict subject compliance with clinical trial protocols. That is, the markers or variables used to predict noncompliance may or may not, in and of themselves, be related to noncompliance. Algorithms may rely on different variables for different subgroups. For example, subgroups can include, but are not limited to, men and women, older or younger subjects, or subjects late or early in the clinical trial. The algorithms may also identify different combinations of variables working only in tandem. Thus, the variables alone may not be directly, simply, or obviously related to noncompliance. The algorithms of the invention may use complex and nonintuitive combinations of predictors to predict subject noncompliance with clinical trial protocols.

The invention also allows context-specific algorithms to maximize predictive utility. That is, different algorithms can be derived for different groups of subjects, such as, for example, subjects with cardiovascular or central nervous system diseases. As a result, the present invention avoids assuming that any given set of variables will be equally predictive of clinical trial noncompliance for all types of conditions or diseases or for all types of subjects.

According to an embodiment of the invention, the present invention also provides for novel quantitative analyses to be incorporated into the development of algorithms to further improve their predictive validity. Specifically, the algorithms can be subject to continuous improvement as more data become available for analysis, either within an individual clinical trial or accumulating across clinical trials.

According to a further embodiment of the invention, algorithms can be used to determine appropriate feedback to both subjects and research staff regarding compliance with the protocol. For example, a computer program can implement decision rules based on these algorithms, and automatically determine appropriate feedback or action by the personnel involved in conducting the clinical trial.

An advantage of the present invention is that, because the algorithms and decisions are based on formal, empirical, and quantitative criteria, they remove the subjective element in these decisions, which in turn minimizes the potential for bias.

The above and other advantages provided by the disclosed invention include provisions for the application of these algorithms within existing and yet to be developed processes for giving systematic feedback to subjects, research sites, and clinical trial sponsors conducting clinical trials using clinical trials.

Once the one or more algorithms of the invention have been derived from analysis of existing data, the algorithms can be translated into specific decision rules, step 130. Decision rules are essentially reformatted algorithms that can be applied to current subject compliance data to determine whether action is needed, step 140. Decision rules may determine a threshold of compliance or a threshold of noncompliance. Optionally, a decision rule may identify a spectrum of noncompliance, from minor noncompliance needing only corrective feedback, to significant noncompliance requiring subject removal from the clinical trial. Decision rules may be based on the specific dependent variable used to derive the algorithm or may be based on one or more differing variables.

For example, a subject who, within the first two days of the clinical trial, does not respond to more than 20% of prompted inquiries and either suspends prompting more than once or indicates he/she is napping more than once may be identified as failing to comply with the research protocol. As another example, subjects who suspend prompting at least twice, and whose total time of such suspension exceeds 2 hours, might be determined to be likely noncompliant, regardless of their overall performance. For purposes of illustration, one sample decision rule may be stated as:

Display noncompliance remediation message to clinical staff if:

$$[.32(\text{ratio of missed random prompts}) + .45(\text{mean number of minutes spent time delaying assessments per day}/100) + .80(\text{mean number of hours spent in sleep each night over past 7 days}/10)] > 1$$

where if noncompliance is determined by this decision rule, an action, such as sending a specific message to the clinical staff is recommended. For example, in the present example, the message "Subject is not following the protocol as required, resulting in substantial missed assessments. Call subject." may be determined to be the appropriate action.

According to an embodiment of the invention, criteria for identifying a subject as noncompliant with the research protocol need not overlap with criteria developed for determining whether to drop a subject from the clinical trial or exclude data related to that subject from the clinical trial results. For example, the decision rule(s) related to dropping a subject from the clinical trial might be based on failed responses to audible prompts rather than on suspension of prompting.

Typically, a decision rule specifies what type of action is required and may provide specific action details. Action types include corrective, affirmative and anti-fraud actions. Action details may include the content of a message to be provided to a subject or to clinical staff.

Decision rules may be translated from algorithms that identify patterns of non-compliance data that are harbingers or leading indicators of later, more serious, non-compliance. This would allow early action to be taken based on these indicators. Such decision rules would typically be in the form of contingencies or conditions based on early compliance indicators.

Optionally, translation of algorithms to decision rules may involve human input or additional factors. For example, balancing the impact of a decision rule

against the focus of the clinical trial may result in an alteration of the decision rule. For example, if subjects' heart rates are being monitored, frequency of prompting or loudness of reminder alerts may be minimized so as not to artificially raise subject heart rates. Also, clinical staff may alter decision rules based on their assessment of external factors outside of the scope of the quantitative analysis. An example may include providing more alerts to clinical staff instead of directly to subjects to provide more interaction between clinical staff and the subjects.

A decision rule may also be used to predict which subjects will fail to complete a clinical trial protocol. Therefore, a decision to rule to drop the subject from the clinical trial, or to work to improve subject performance, can be made at an early time. By providing those conducting a clinical trial with early feedback regarding subject noncompliance with a research protocol, the present invention improves clinical trial data quality and may potentially save both time and money by either improving the compliance of potentially noncompliant subjects or excluding unimprovable noncompliant subjects early in a clinical trial.

The generation of a fraud detection algorithm can take many forms. The psychometric properties of the scale itself could be used to identify potentially fraudulent responses. For example, according to one embodiment of the invention, item response theory uses known properties of individual items within an assessment to estimate the probability that an observed pattern of responses is valid. Therefore, a subject answering yes to the question "My headaches are completely debilitating" has a low probability of also answering yes to the question "My headaches are a minor inconvenience" such that observing this pattern of responses could be indicative of fraud.

According to a further embodiment of the invention, the detection of fraudulent or invalid entries in subject-supplied data may be performed similarly to the methods described herein. For example, the analysis could be based on statistical properties of the responses themselves. Thus, as an example, analysis might indicate that when the standard deviation across subject responses on a particular questionnaire are less than 1.0, fraudulent or invalid completion is highly likely.

The content of subjects' responses could optionally be used as a source of data for the fraud detection algorithms if the responses are invariant or relatively invariant.

For example, a subject answering 'yes' to all questions, even when the logical content of the questions would suggest some alternating pattern of appropriate responses.

Analysis of fraud could also be based on particular combinations of responses. Thus, subjects who answered that they took pain medication five or more times daily, but who elsewhere indicated either that pain severity was 4, on a scale of 1 to 10, or that pain frequency was 'infrequent' or 'rare', might be flagged as cases likely to be invalid. The response patterns determined to represent potentially fraudulent data need not be logically inconsistent or intuitively invalid. Rather, they are determined to represent potentially fraudulent data based on statistical analysis comparing valid and invalid response profiles. Therefore, questions posed to subjects can be tailored to provide opportunities for the subject to contradict, or appear in disagreement with, responses to earlier questions.

In an alternative embodiment, the posing of questions providing opportunities to contradict earlier responses can be interactive. For example, further questions providing opportunities to contradict earlier responses can be posed only if a response to a question appears unusual or if a decision rule indicates earlier indications of potential fraud.

As a further example, the time required for a subject to respond to items could be the foundation for the generation of fraud detection algorithms. For example, evaluability data could be used to estimate the mean length of time subjects take to respond to certain items. In such an example, response latencies less than two standard deviations below those norms could be the basis of identifying the responses as potentially fraudulent. For example, if a question contains 25 words and subjects take an average of 8 seconds to answer the question, responses of less than 1 second could be identified as potentially fraudulent.

Alternatively, the portable electronic device could capture certain ecological data such as temperature or airborne particles, or physiological data, such as concurrent heart rate, suggestive of a particular location, subjective, or physical state, which is inconsistent with the subject's responses, suggesting possible fraud.

In an alternative embodiment of the invention, subjects can be included in the clinical trial for the purpose of providing fraudulent data. For example, in a group of 100 subjects, 20 subjects may be asked to provide fraudulent data. By having such fraudulent data among data provided by the subjects, the quantitative analysis of the

present invention can be used to ensure the resulting algorithms and decision rules detect the known fraudulent entries. In the event other subjects are also fraudulently recording data without the knowledge of the clinical staff, the algorithms and decision rules will likely also detect such unknown fraudulent activity.

- 5 Each of the above variations for detection of fraud can be used according to various embodiment of the present invention individually, sequentially or in combination.

- According to a preferred embodiment of the invention, the system of the invention for automated processing of data collected via a portable electronic device is provided. In this embodiment, the portable electronic device or system is designed to prompt a subject for information and/or collect information as recorded by the subject without prompting. Preferably, each subject in the clinical trial is provided with a portable electronic device. The portable electronic device is preferably used to collect compliance-relevant variables, such as the number of data entry episodes, missed data entry occasions (e.g., instances where the portable electronic device prompts for data, but the subject fails to respond). A database of these variables is preferably processed according to the decision rules to guide the actions of the portable electronic device as described in detail in the copending patent application titled "System for Clinical Trial Subject Compliance", Attorney Docket No. IVQ-002.

- 20 The portable electronic device is also preferably adapted to communicate with another computer to allow the clinical staff to consolidate the data from all subjects in the clinical trial into one location for review or processing. Preferably, the portable electronic device will also be adapted to communicate with at least one other computer via a wireless connection or via a wired connection, including the use of a modem and/or a network, such as a LAN or the Internet. For example, by the use of the Internet or a dial-up modem connection, a subject may submit information from the portable electronic device to the clinical staff from the subject's home.

- In another embodiment, a portable electronic device or a computer is adapted to communicate with clinical trial equipment used for measuring, monitoring, controlling or recording data or a process of the clinical trial. Examples of such processes include administration of medication or monitoring of heart rates. The portable electronic device or a computer preferably automatically records desired data for incorporation in the clinical trial data or compliance data.

In another embodiment of the invention, a paper form, such as a case report form, can be used by the subject to record data. The data can then be entered into a database by the use of a portable electronic device or other computer at an appropriate time. Examples of case report forms include hand-written forms and forms that allow for machine readable marks to be made, enabling automated scanning of the case report forms during entry of the data into a computer.

In an alternative embodiment of the present invention, the methods of the present invention may be incorporated in instructions recorded on a medium suitable for use in an electronic device, such as a computer, computer network server or a portable electronic device. The medium can include, for example, a hard disk, RAM medium, diskette, CD-ROM or other optical or magnetic storage medium. The instructions can optionally be stored on a server that can be remote from the subject or clinical staff member.

According to an embodiment of the invention, the server can provide data to be displayed. Data may be displayed at the server itself or be transmitted to another location, such as via hardwired or wireless access to the server, including a LAN or the Internet. The data can be processed to provide a graphical display to interested parties. Examples of those who may be interested in viewing the graphical representation of the compliance data include a site coordinator (who may be interacting with the subject), a clinical research organization (who may be responsible for study execution across a number of research locations), other agencies interested in the collection of the data, or the sponsor of the research.

According to another embodiment of the invention, the server can provide ongoing aggregation of data across subjects to speed the time required to combine, clean, and make available final data.

These examples are meant to be illustrative and not limiting. The present invention has been described by way of example, and modifications and variations of the exemplary embodiments will suggest themselves to skilled artisans in this field without departing from the spirit of the invention. Features and characteristics of the above-described embodiments may be used in combination. The preferred embodiments are merely illustrative and should not be considered restrictive in any way. The scope of the invention is to be measured by the appended claims, rather than

the preceding description, and all variations and equivalents that fall within the range of the claims are intended to be embraced therein.

Having described the invention, what is claimed as new and protected by Letters Patent is:

1. A method of determining a value of a function $f(x)$ at a point x by using a series of values $f(x_1), f(x_2), \dots, f(x_n)$ and a set of weights w_1, w_2, \dots, w_n such that $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0$ for all i . The method includes the steps of: (a) determining the values $f(x_i)$ and the weights w_i ; (b) calculating the weighted sum $\sum_{i=1}^n w_i f(x_i)$; and (c) outputting the result as the value of $f(x)$.